An Efficient Image Self-recovery and Tamper Detection Using Fragile Watermarking

Sajjad Dadkhah[1(✉)], Azizah Abd Manaf[2], and Somayeh Sadeghi[3]

[1] Faculty of Computing, Universiti Teknologi Malaysia,
54100 Kuala Lumpur, Malaysia
dsajjad2@live.utm.my
[2] Advanced Informatics School, Universiti Teknologi Malaysia,
54100 Kuala Lumpur, Malaysia
azizah07@ic.utm.my
[3] Faculty of Computer Science and Information Technology, University of Malaya,
Kuala Lumpur, Malaysia
ssomayeh@siswa.um.edu.my

Abstract. Fragile watermarking is one of the most effective approaches to insure the integrity of digital images. In this paper, an efficient self-recovery and tamper localization scheme using fragile watermarking is proposed. The proposed method generates 12-bit tamper detection data and 20-bit self-recovery data for each 4×4 block. The generated tamper detection and self-recovery features are encrypted by utilizing user secrete key. A random block mapping scheme is used to embed the encrypted block features into its mapping block. The proposed two-level tamper detection creates high capacity for tamper detection data which improves the security and tamper localization. The performance of the proposed scheme and its robustness against famous security attacks is analyzed. The experimental results demonstrate the high efficiency of the proposed scheme in terms of tamper detection rate, tamper localization and self-recovery. This method is robust against security attacks such as collage attack and constant average attack.

Keywords: Tamper detection · Tamper localization · Self-recovery · Fragile watermarking · Image security

1 Introduction

The integrity and authenticity of the digital images can be assured by utilizing the tamper detection algorithms that use watermarking techniques . Fragile watermarking is one of the most effective methods to be used for tamper detection and tamper localization [1]. In recent years, various fragile watermarking schemes for tamper detection and self-recovery have been proposed [2–7]. Generally, the digital images that are watermarked by these schemes are partitioned into non-overlapping blocks of pixels. The generated watermark feature for tamper detection and recovery is embedded into blocks with different locations which

A. Campilho and M. Kamel (Eds.): ICIAR 2014, Part I, LNCS 8814, pp. 504–513, 2014.
DOI: 10.1007/978-3-319-11758-4_55

makes them robust against certain malaciouse attacks. However, these tamper detection and self-recovery methods struggle with a few more problems.

1. **Lack of tamper localization precision**
   Tamper detection schemes with self-recovery capability embed the generated watermark information into blocks with different locations. Therefore, if a mapping block that contains watermark information of a different block is destroyed, two blocks will be detected as tampered. To address this issue, Lin et al. [2] proposed a hierarchical tamper detection algorithm with self-recovery capability. The tamper detection data generated by Lin's algorithm is embedded as watermark payload into the same block and self-recovery data is embedded into a different block. The embedding procedure proposed by Lin has been adopted by several researchers [2,3,5–9].
2. **Insufficient embedding capacity**
   Because of the insufficient embedding capacity, certain constant information such as average intensity of the block or certain features of discrete cosine transform (DCT) coefficients are used by the fragile tamper detection schemes. These methods [2,3,8,9] are incapable of detecting tampering attacks that do not modify their designated feature. The dual watermarking algorithm proposed by Lee and Lin [3] suffer from this problem. Their proposed algorithm offers a second chance of recovery survival, but in contrast any modification that alters bits in 5 MSB (most significant bit) or higher positions cannot be detected by their algorithm.
3. **Lack of security for embedding procedure**
   The blockwise dependency ensures the robustness of self-recovery algorithm [6–8,10] against common security attacks such as vector quantization (VQ) [11]and collage attack [12]. However, these schemes are vulnerable against tampering attacks which use the same block mapping scheme to locate the generated watermark data. Several tamper detection algorithms suffer from lack of sufficient security measurement such as secret key for encrypting in the embedding procedure.

To resolve the tamper localization and security problems that are mentioned above, this paper proposes an efficient tamper detection and self-recovery algorithm based on fragile watermarking with following characteristics:

1. Generate 12-bit tamper detection based on block binary feature and 20-bit average intensity for self-recovery.
2. Generate encrypted block-mapping algorithm based on security key and encrypt the inserted information of each block 4×4 pixels.
3. Apply a second-level of tamper detection by generating new 20-bit tamper detection keys, which are embedded in different block to eliminate security attacks, such as a VQ counterfeiting and collage attack.

The remainder of this paper is organized as follows. In section 2, the proposed fragile watermarking for tamper detection and self-recovery is described. Section 3 presents the performance analysis and experimental results. The paper's conclusions are presented in section 4.

2 The Proposed Algorithm

The proposed fragile tamper detection and self-recovery scheme is explained in three phases : watermark generation and embedding, tamper detection and localization, self-recovery.

2.1 Watermarking Scheme

The proposed watermarking scheme is encrypted by using a user secret key. For digital images which have more than one color space such as RGB images, all color channels (red, green and blue) are watermarked by the proposed algorithm. The proposed watermarking procedure consists of the following steps:

Step 1. Preprocessing.The original image O is divided into M 4×4 blocks F_i, and F_i is divided to four 2×2 blocks G_i. The two least significant bit of each pixel is converted to zero.

Step 2. Tamper Bit Generation. Each 4×4 block F_i is decomposed as $F_i = R_i^+||C_i^+$, where R_i^+ is the addition result of pixels in each row, and C_i^+ is the addition result of pixels in each column of 4×4 blocks, and $||$ is bitwise concatenation. As following equations illustrate, the $12_{bit}TDK$ is the tamper detection key for each block of 4×4 pixels.

$$12_{bit}TDK = 8_{bit}(A_{avg})||4_{bit}(F_n) \tag{1}$$

$$K_i = mod(R_i^+, 2) + mod(C_i^+, 2), i = (1, ..., 4) \tag{2}$$

$$4_{bit}(F_n) = \begin{cases} 1 \;\; if \;\; mod(CO(K_n, 1), 2) = 0, (n = 1, ..., 4) \\ 0 \;\; if \;\; mod(CO(K_n, 1), 2) = 1, (n = 1, ..., 4) \end{cases} \tag{3}$$

$$8_{bit}(A_{avg}) = mod(Av_{F_n}/2^{j-1}, 2), j = (1, ..., 8) \tag{4}$$

where K_i is the total binary summation of each row pixel addition R_i^+ with column pixel addition C_i^+ for the same i value. Moreover, for each 4×4 block , four R_i^+ and four C_i^+ value are generated, and as i in equation 2 illustrated, four value of K_i are generated for each 4×4 block. The $CO(K_n, 1)$ in equation 3 presents the total number of 1's in binary form of $K_n(n = 1, ..., 4)$. As illustrated in equation 3, four $F_n(n = 1, ..., 4)$ value is generated for each 4×4 block. However, if the value of $CO(K_n, 1)$ is even, F_n will be set to 1, otherwise, F_n will be set to 0. Av_{F_n} in equation 4 present the average intensity of 4×4 blocks, and $8_{bit}(A_{avg})$ is the 8-bit binary form of average intensity.

Step 3. Self-recovery Bit Generation. The self-recovery data generated by the proposed scheme, is 20-bit key RCK, which consists of the five most significant bits (5MSB) of each average intensity of 2×2 blocks G_i.($20_{bit}(RCK) = G_i^1||G_i^2||G_i^3||G_i^4$)

Step 4. Encryption. The 32-bit watermark data which is decomposed as $32Wt_{bit} = 12_{bit}(TDK)||20_{bit}(RCK)$, is encrypted with the following equation,

$$E(Wt_{bit}) = 32Wt_{bit} \oplus K_s \tag{5}$$

where K_s is user secret key and $\oplus$ is the exclusive or (XOR). The user key K_s will be obtained by user in the beginning of each watermarking and tamper detection procedure. However, as equation 5 illustrated, the generated 32-bit watermark data $32Wt_{bit}$ for each 4×4 block will be encrypted by user key K_s which is only known to user. The optimization conducted in this step secures the proposed scheme against famous security attacks such as four-scanning attack.

Step 5. Embedding and Block-mapping. The generated $12_{bit}(TDK)$ is embedded into the least significant bit of each pixel inside $4\times$ 4 block F_i. However, as Fig.1 demonstrates, the $20_{bit}(RCK)$ is embedded into first and second least significant bit of a random selected mapping block. Fig.1 shows that the proposed block-mapping algorithm selects a random block with the most distanced location from the original block.

```
1:  b = (Block.no ÷ 2) + (width ÷ 4)
2:  d = (Block.no ÷ 2) + 1
3:  Process
4:    for all (rand(i) ∈ b) ⊆ d do
5:      if i ≤ d then
6:        for all ee ∈ {1 − 20} do
7:          if ee < 16 then
8:            2LSB {A_i^block(1, ee)} ← ∀ Get. Bit_i^TDK(1, ee)
9:          else
10:           1LSB {A_i^block(1, ee)} ← ∀ Get. Bit_i^RCK(1, ee)
11:         end if
12:       end for
13:       b ⇐ b + (width ÷ 4)
14:       d ⇐ b − (width ÷ 4) + 1
15:       decrement i by one
16:     end if
17:   end for
18: end Process
```

Fig. 1. Embedding and block-mapping algorithm

2.2 Tamper Detection and Localization Algorithm

The proposed tamper detection and tamper localization algorithm locates the manipulated regions of the watermarked image $Wt_i, (i = 1, ..., n)$, and marks the suspicious block as either valid $v_i = 1$ or invalid $v_i = 0$. However, the optimization proposed in this section creates a blockwise dependency which secures the proposed scheme against tamper attack such as the collage attack. The details of the proposed tamper detection method are explained as follows.

Step 1. Block partitioning. The same procedure as Step 1 in Section 2.1 will be conducted.

Step 2. Retrieve watermark. Extract the $12_{bit}(TDK)$ from each 4×4 block F_i and $20_{bit}(RCK)$ from mapping blocks. The mapping block location is determined with the same procedure presented in Step 5 of Section 2.1.

Step 3. Tamper detection. In this step, the 12-bit tamper detection data TDK^n and 20-bit self-recovery RCK^n is reconstructed with the same procedure presented in Steps 1 and 2 of Section 2.1. The extracted detection key TDK and new generated key TDK^n are compared as:

$$v_1^i = \begin{cases} 1 \;\; if\, 12_{bit}(TDK) = 12_{bit}(TDK^n) \\ \\ 0 \;\; if\, 12_{bit}(TDK) \neq 12_{bit}(TDK^n) \end{cases} \tag{6}$$

where the new generated self-recovery information RCK^n will be used for tamper detection purpose and the tampered blocks will be marked as tamper $v_1^i = 0$. Moreover, the following expressions secure the proposed scheme against malicious attack such as collage tampering.

$$v_2^i = \begin{cases} 1 \;\; if\, v_1^i = 1\, and\, 20_{bit}(RCK) = 20_{bit}(RCK^n) \\ \\ 0 \;\; if\, v_1^i = 1\, and\, 20_{bit}(RCK) \neq 20_{bit}(RCK^n) \end{cases} \tag{7}$$

Step 4. Tamper localization. The 4×4 blocks F_i with $v_1^i = 0$ or $v_2^i = 0$ will be presented as tampered regions.

2.3 Recovery Algorithm

The proposed tamper localization scheme distinguishes the tampered blocks by marking them as valid or invalid. However, to identify the blocks that need to be recovered, the array $N_i^R, i = (1, .., N)$ is generated by the following expression:

$$N_i^R = \begin{cases} 1 \;\; if\, v_1^i = 1 \,\&\, v_2^i = 1 \\ \\ 0 \;\; if\, v_1^i = 0 \,\&\, v_2^i = 0 \end{cases} \tag{8}$$

The 20-bit self-recovery features RCK_i^n is extracted from its mapping block by Step 2 of Section 2.2. The 20-bit RCK_i^n consist of four 5-bit values which are extracted from 5MSB of the average intensity of the 2×2 blocks G_i^n . Moreover, if $N_i^R = 1$ the tampered block F_i^R will be recovered by substituting the extracted self-recovery information RCK_i^n with the destroyed pixels in F_i^R. Since the proposed self-recovery information is constructed based on 2×2 blocks, the quality of the recovered image will be improved after self-recovery procedure.

3 Experimental Results

To evaluate the performance of the proposed tamper detection and self-recovery algorithm, two distinct measurements are introduced in this section. Generally,

different quality measurements such as signal to noise ratio (SNR), peak signal to noise ratio (PSNR), Mean square error (MSE) and Watson distance (WD) will be used to evaluate the quality of watermarked image and self-recovery scheme. In this paper MSE and PSNR are used as follows :

$$\text{MSE} = \frac{1}{mn} \sum_{j=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2, \tag{9}$$

$$\text{PSNR} = 10.Log_{10}(\frac{b}{\text{MSE}}), \tag{10}$$

where b is the square of the maximum value of the signal, and m × n denotes the dimensions of the monochrome image of I and K. In this paper, i.e., image intensity is of 8 bits format, so b= 255^2. The second performance measurement which is used for evaluation, is tamper detection rate T_{dt}.

$$T_{dt} = (1 - (FP_r + FN_r)/(1 + P)) \times 100 \tag{11}$$

where FP_r is false positive rate and FN_r is false negative rate and P is the number of regions which have been manipulated.

3.1 Malicious Tampering Attacks

To evaluate the security robustness of the proposed scheme, several malicious attacks such as collage attack, constant-average attack(CAA) [13] and VQ attack are examined. In addition, several general tampering such as deletion attack and drawing attack are also examined. Fig. 2 shows the visual experimental results of the proposed scheme against several malicious tampering attacks. The Blond, Color Lena, Pirate and Barbara images with size of 512 × 512 are selected. The Magazine and Soldier images with size of 512 × 512 and 400 × 290 are collected from [14]. All test images in this experiment are watermarked with the same secret key. Thus, it is assumed that the attacker has knowledge about the contents of the secret key and proposed scheme structure.

As seen in Fig.2a,the watermarked Blond, color Lena and Pirate images generated by the proposed watermarking algorithm, 2e and 2i, has the PSNR of 43.21, 43.87 and 43.70 dB, respectively. Fig.2b is the tampered image with tamper ratio less than 20 %. As shown in Fig. 2d, the recovered image produced by the proposed algorithm, has the PSNR of 38.42 dB. Fig. 2f, represents three type of distinct tampering attacks:(1) Square deletion and rectangle deletion, (2) VQ tampering attack: copy woman and nightstand lamp from watermarked room image and place it on different spatial locations inside Lena image. The room image has been watermarked by the proposed scheme with the same secret key used for watermarked Lena image, and (3) Writing tampering: some letters "UTM" with red color is placed in Lena image. The tamper ratio for the multi region tampering attack in 2f is more than 30 %. Fig.2k is the collage tampered Pirate image, in which more than 40% of the watermarked Barbara image is copied and pasted into watermarked Pirate. Moreover, in this attack, the spatial

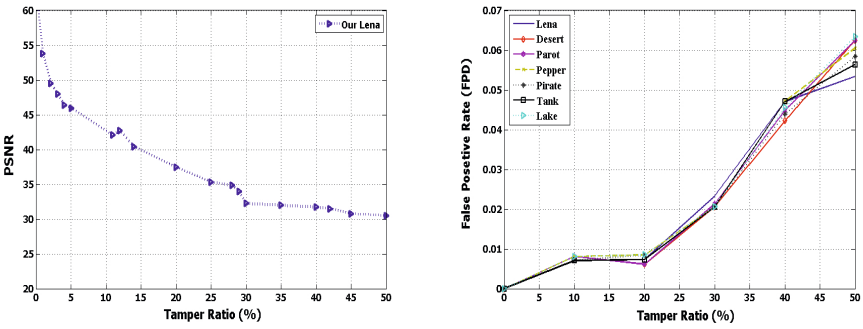
Fig. 2. Malicious Tampering Attacks. (a) Watermarked Blond ,(b) General tampering, (c) Tampered located, (d) Recovered image (PSNR= 38.42 dB), (e) Watermarked Color Lena(PSNR= 43.87 dB), (f) Multi Tampering , (g), Tampered located, (h) Recovered image (PSNR= 34.22 dB),(i) watermarked pirate, (j) Collage attack , (k) Attack detected ,(l) Recovered image (PSNR= 33.44 dB),(m) Watermarked magazine, (n) Collage attack 50 % , (o) 50 % CA attack detected ,(p) Recovered image ,(q) Watermarked soldiers, (r) CAA and VQ attack , (s) Attack detected ,(t) Recovered image

locations of the copied watermarked Barbara image is preserved in the watermarked pirate image.

As illustrated in Fig. 2c, 2g and 2k, the proposed tamper detection algorithm accurately located all the tampered regions, the black color regions are the authentic part of the tested images. The proposed tamper detection algorithm is very efficient in indicating the tampered and original 4×4 blocks. As shown in , Fig. 2d, 2h and 2l, the self-recovery scheme completely recovered all the tampered regions. The recovered Blond,Lena and pirate images, have the PSNR of 38.42, 34.22 and 33.44 dB, respectively. As Fig.2m, 2n, 2o and 2p demonstrated, the 50 % CA attack is completely detected and recovered. Different pixels of block 4×4 in Fig.2r are modified with CAA attack and multiple regions of the images are replaced with VQ attack. As 2s and 2t illustrated, all the tampered areas are completely detected and recovered. As the experimental results demonstrate in Fig. 2, the proposed scheme is able to efficiently locate and recover different types of tampering such as general tampering, VQ attack, Collage Attack and multi region tampering attack with a satisfying PSNR values. However, the proposed self-recovery algorithm is efficient in recovering the tampered region because of the selected recovery block size, which is 2×2 pixels.

3.2 Performance Analysis and Evaluations

In this section, the performance of the proposed tamper detection and self-recovery scheme is analyzed. Fig. 3 shows the PSNR and Fp_r value of the recovered images by the proposed scheme, for different tamper ratios. As seen in Fig. 3a , the PSNR of the recovered image for tampering attack with tamper ratio less than 10 % is fairly high. However, Fig.3a shows that the proposed self-recovery algorithm achieve the satisfying PSNR value of 31.00 dB for tamper ratio of 50 %.



(a) (b)

Fig. 3. Performance Analysis.(a)Recovered image PSNR (b) FP_r values

Fig. 3b shows that the false positive rate of the proposed algorithm is slightly increased with increase of tamper ratio. However, with the increase of FP_r, tamper detection rate T_{dt} will be degraded. It can be seen from Fig. 3b, that the FP_r of the proposed tamper detection algorithm is always less than 0.1 for different tamper ratios, and the FP_r of the different images are almost the same. The similarity of the false positive rate value for different images shows that, the images with different complexity does not have any degrading effect on the performance of the proposed tamper localization scheme. Moreover, after examining several digital images, the value of FP_r for different tamper ratio remained less than 0.1, and the value of false negative rate FN_r is very low. The tamper detection rate T_{dt} of the proposed scheme, which is obtained by equation 11, is higher than 99 % for general tampering attacks.

Table 1 presents the performance comparison of the proposed tamper detection and self-recovery algorithm against different malicious attacks such as collage attack. The 512×512 Lena image is used for performance analyses in Table 1. Table 1 shows the good recovery quality of Lee' s algorithm [3] for tamper ratio higher than 30 %, but his algorithm is not robust against any of malicious attacks. Patra [5] and Tong's [6] scheme generate PSNR lower than 31.00 dB for recovered image and their algorithms are not robust against all the malicious attacks mentioned in Table 1. However, as Table 1 demonstrates, the proposed method outperform other algorithms in security robustness and self-recovery.

Table 1. Performance Comparison of Self-recovery and Security robustness

Methods	Watermark PSNR (dB)	Recovered 30% tamper PSNR (dB)	Collage attack	VQ	CAA
Lee [3]	40.68	36.39	No	No	No
Patra [5]	43.94	31.41	Yes	Yes	No
Tong [6]	40.73	27.30	No	Yes	Yes
Proposed	43.87	37.23	Yes	Yes	Yes

4 Conclusions

In this paper, an efficient tamper detection and self-recovery scheme using fragile watermarking is proposed. The proposed tamper detection scheme generates 12-bit tamper detection based on block binary feature and 20-bit average intensity for self-recovery. The proposed tamper localization algorithm accurately locates the tampered blocks of size 4×4 pixels and the self-recovery scheme recovers the four blocks of size 2×2 pixels within the tampered block. The proposed random block-mapping algorithm creates robustness against security attacks such as collage attack CAA and VQ attack. However, the performance analysis and experimental results clearly demonstrate the efficiency of the proposed scheme in terms of tamper localization, security robustness and recovery quality. Future research include utilizing block-neighboring characteristic to recover the tampered blocks whose recovery information is destroyed.

Acknowledgments. The authors would like to thank Universiti Teknologi Malaysia for its educational and financial support.This work is funded by the FRGS grant under Vote No. 4L043 which is supported by Universiti Teknologi Malaysia (UTM) and the Ministry of Higher Education (MOHE).

References

1. Dadkhah, S., Manaf, A.A., Sadeghi, S.: Efficient digital image authentication and tamper localization technique using 3lsb watermarking. International Journal of Computer Science Issues (IJCSI) **9** (2012)
2. Lin, P.L., Hsieh, C.-K., Huang, P.-W.: A hierarchical digital watermarking method for image tamper detection and recovery. Pattern Recognition **38**(12), 2519–2529 (2005)
3. Lee, T.-Y., Lin, S.D.: Dual watermark for image tamper detection and recovery. Pattern Recognition **41**(11), 3497–3506 (2008)
4. Yang, C.-W., Shen, J.-J.: Recover the tampered image based on vq indexing. Signal Processing **90**(1), 331–343 (2010)
5. Patra, B., Patra, J.C.: Crt-based fragile self-recovery watermarking scheme for image authentication and recovery. In: IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 430–435 (2012)
6. Tong, X., Liu, Y., Zhang, M., Chen, Y.: A novel chaos-based fragile watermarking for image tampering detection and self-recovery. Signal Processing: Image Communication **28**(3), 301–308 (2013)
7. Chang, C.-C., Chen, K.-N., Lee, C.-F., Liu, L.-J.: A secure fragile watermarking scheme based on chaos-and-hamming code. Journal of Systems and Software **84**(9), 1462–1470 (2011)
8. Zhu, X., Ho, A.T., Marziliano, P.: A new semi-fragile image watermarking with robust tampering restoration using irregular sampling. Signal Processing: Image Communication **22**(5), 515–528 (2007)
9. Wang, L.-J., Syue, M.-Y.: A wavelet-based multipurpose watermarking for image authentication and recovery. International Journal of Communications **2**(4) (2013)
10. Qian, Z., Feng, G., Zhang, X., Wang, S.: Image self-embedding with high quality restoration capability. Digital Signal Processing **21**(2), 278–286 (2011)
11. Holliman, M., Memon, N.: Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. IEEE Transactions on Image Processing **9**(3), 432–441 (2000)
12. Fridrich, J., Goljan, M., Memon, N.: Cryptanalysis of the yeung-mintzer fragile watermarking technique. Journal of Electronic Imaging **11**(2), 262–274 (2002)
13. Chang, C.-C., Fan, Y.-H., Tai, W.-L.: Four-scanning attack on hierarchical digital watermarking method for image tamper detection and recovery. Pattern Recognition **41**(2), 654–661 (2008)
14. Ng, T.-T., Chang, S.-F., Hsu, J., Pepeljugoski, M.: Columbia photographic images and photo realistic computer graphics dataset, Columbia Univ., New York, ADVENT Tech. Rep., 205–2004 (2005)